Enabling Preservation by means of Open Source

dave rice
@dericed
#fosdem
2015-01-31


JVC
Laird
Bellari
6124

Demolition of Sony Magnetic Tape Plant - Dothan, Alabama

from **Steve Frank Films** PLUS 3 years ago NOT YET RATED

The demolition of the Sony Magnetic Tape Plant located in Dothan, Alabama


SONY
Avid
& STORAGE
BRUNO
ATW
CANAPE
FRIEND
FOOT

- Preservation of audiovisual material like the preservation of a species requires creation of copies
- Preservation copies are generated to deter obsolescence risks.
- Archivists must negotiate challenges from increased collections size and constrained resources.
- One-on-one human-object interaction in preservation must be prioritized and selective rather than mandatory.
- The practicality of conservation-only approaches is the weakened. Archives must be run differently than time capsules.

"How to move from one-to-one workflows to many-to-less workflows?"

"How can we preserve more than we could preserve before?"

"Is technology a replacement, a trusted co-worker? Is this a healthy relationship?

BACKBLAZE

English

Home | How it Works | Buy Now | About Us | Partners | Help Me | Sign In to Restore

Relax... your data will be backed up in 259,260 days.

FAIL

- Use your computer normally – backups happen in the background
- After the initial backup, incremental backups will be fast
- Your backup need not finish during your trial – you can test a restore anytime

Test your connection speed to Backblaze datacenter

Reformat to What?

Reformat to What?

Sustainability Factors
Disclosure
Adoption
Transparency
Self-Documentation
Metadata Capabilities
Impact of Patents

Reformat to What?

<u>Sustainability Factors</u>
Disclosure
Adoption
Transparency
Self-Documentation
Metadata Capabilities
Impact of Patents

<u>Cost Factors</u>
Cost of Software
Cost of Hardware
Storage Cost
Network Cost

Reformat to What?

<u>Sustainability Factors</u>
Disclosure
Adoption
Transparency
Self-Documentation
Metadata Capabilities
Impact of Patents

<u>Cost Factors</u>
Cost of Software
Cost of Hardware
Storage Cost
Network Cost

<u>Implementation Factors</u>
Difficulty to Implement
Complexity of Toolsets
Availability to Tools: For transcoding, metadata, qc, etc
Ease of Format Identification / Validation

LOC <3 JPEG2000!!!

Photo Credit: CC-BY-NC Sarah Ross Photography

SAMMA Inputs and Outputs

Tape Formats Supported

- Betacam, SP, SX, Digital Betacam
- U-matic small and large cassettes
- VHS, S-VHS, MII, D3

Output Formats

- Video Tape, LTO Tape, DLT Tape
- SDI, Embedded audio or AES
- MPEG-2, H.264, IMX files
- Motion JPEG 2000 (lossless)
- Windows Media, QuickTime

SAMMA SYSTEMS


You did it! Congratulations! World's best cup of coffee!
Great job, everybody! It's great to meet you.

Jimi Jones, Digital Audiovisual Formats Specialist, Library of Congress

"As dissatisfying as it is, there is as yet no good, pat answer to digital video preservation, largely because the digital preservation world is so emergent. Efforts like FADGI's [Federal Agencies Digitization Guidelines Initiative] MXF/JPEG 2000 work may help the digital preservation community to someday have a better answer to the 'which digital video format' question. For the moment the 'right answer' is actually more of a 'best set of questions to ask.'"

http://blogs.loc.gov/digitalpreservation/2011/07/whither-digital-video-preservation/

Uncompressed audiovisual streams in AVI, MXF, MOV ?

The City of Vancouver Archives <3 Archivematica!!!

Photo Credit: CC-BY-NC-SA flickr:qousqous

My groups
Home
Starred

Announcements
Google Groups ...
Recently viewed
ffmbc-discuss
archivematica
ArchivesSpace
serious_video_a...
Recent searches
jpeg2000 (in arc...
ffmpeg (in archiv...
video (in archive...
ffv1 (in perian-di...
Recently posted to
archivematica

Favorites
Click on a group's star icon to add it to your favorites

Encoding JPEG2000-MXF using open-source tools

3 posts by 3 authors in archivematica

Misty De Meo 3/22/11 ☆ Post reply

I've been doing some experimenting with open-source tools to create JPEG2000-encoded MXF video files. I know this is something other people have been having trouble with, so I thought I'd share my findings.

I've been working on Mac OS X, but the tools should also be compatible with Linux and might work in Windows. I've found a few different paths to doing this. There isn't necessarily an ideal start-to-finish workflow yet, but the tools seem to be there to allow one to be scripted or programmed.

Method one: OpenDCP

- FFmpeg
- OpenJPEG 1.4 (http://www.openjpeg.org/)
- OpenDCP (http://code.google.com/p/opendcp/)

This workflow is designed around digital cinema package (DCP) production. Most of the open-source tools are based around DCP since they're being designed by indie filmmakers.

OpenDCP is a utility for creating DCPs. The AS-DCP library (http://www.cinecert.com/asdcplib/) and the mxflib library (http://sourceforge.net/projects/mxflib/) it's based on are possibilities for someone with more programming skills than me to build an archives-targeted tool.

The disadvantage to this route is that DCPs use separate files for audio and video, accompanied by an XML file describing the pair, which might make playback less convenient than a single AV MXF. The available tools

MKV / FFV1 / LPCM ?

ffv1 implementations: Archivematica

Media type preservation plans [edit]

Media type	File formats	Preservation format(s)	Access format(s)	Normalization tool
Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Email	PST	MBOX	MBOX	readpst
Email	Maildir**	Original format	MBOX	md2mb.py
Office Open XML	DOCX, PPTX, XLSX	Original format	PDF for PPTX	OpenOffice
Plain text	TXT	Original format	Original format	None
Portable Document Format	PDF	PDF/A	Original format	Ghostscript
Presentation files	PPT	Original format	PDF	OpenOffice
Raster images	BMP, GIF, JPG, JP2*, PCT, PNG*, PSD, TIFF, TGA	Uncompressed TIFF	JPEG	ImageMagick
Raw camera files/Digital Negative format**	3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F	Original format	JPEG	ImageMagick/UFRaw
Spreadsheets	XLS	Original format	Original format	None
Vector images	AI, EPS, SVG	SVG	PDF	Inkscape
Video	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV	FFV1/LPCM in MKV	MPEG-1	FFmpeg

ffv1 implementations: Österreichische Mediathek

österreichischemediathek
audiovisuelles archiv · technisches museum wien

DVA Profession

Screenshot of the DVA-Profession web gui

Video Digitization

DVA-Profession is a complete solution for digitizing video for archival purposes. It manages the whole workflow, ranging from digitization to analysis, generating preview images and a preview video (MPEG), manual quality control, documentation of all process metadata and the final deposition of the files on a digital mass-storage. All steps of the workflow are designed and optimized for an economic operation and preparation for long-term archiving (for further information see "documentation"). This product is available under a Free Software License (GPLv3 - GNU General Public License) and can be downloaded here. Due to international participation, this site is kept in English.

Videodigitalisierung

DVA-Profession ist eine Gesamtlösung für die Digitalisierung von Video-Material für den Archivgebrauch. Sie verwaltet den gesamten Workflow von der Digitalisierung über Analyse, Erstellen von Vorschaubildern und einer Sichtungskopie (MPEG), manueller Qualitätsüberprüfung, Dokumentation aller anfallenden Metadaten bis hin zum Ablegen in einen digitalen Massenspeicher. Die Arbeitsschritte sind auf ökonomischen Betrieb und Vorbereitung für die Langzeit-Archivierung hin entworfen und optimiert (genauere Beschreibung unter "documentation"). Dieses Produkt steht unter einer Freien Software Lizenz (GPLv3 - GNU General Public License) und kann hier herunter geladen werden.
Diese Seite wird wegen internationaler Beteiligung auf Englisch geführt.

ffv1 version 3

adds multithreaded encoding / decoding

adds mandatory embedded frame crc

self-descriptive

documentation efforts

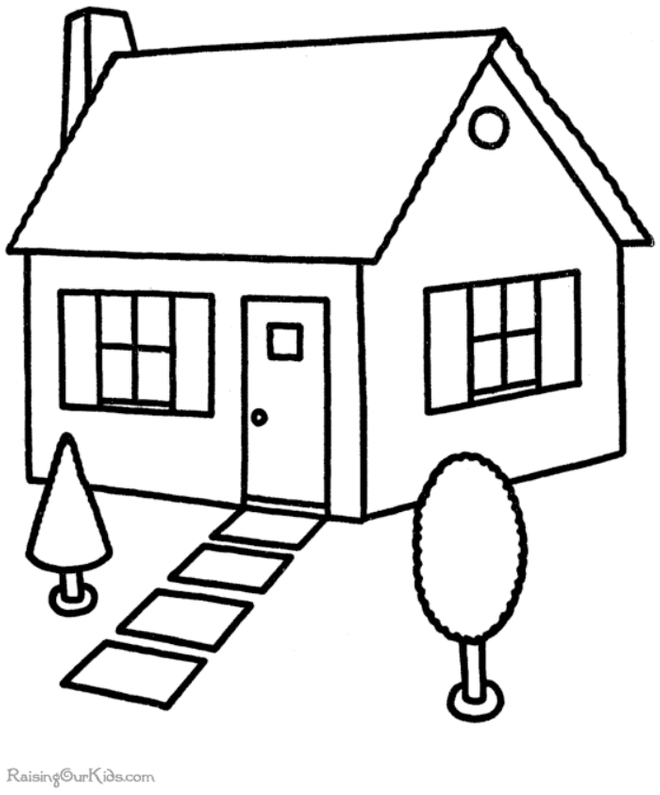
Let's Put Checksums inside Audiovisual Media

MP3
MPEG2

FLAC
FFV1 1.3

Checksum verification analogy

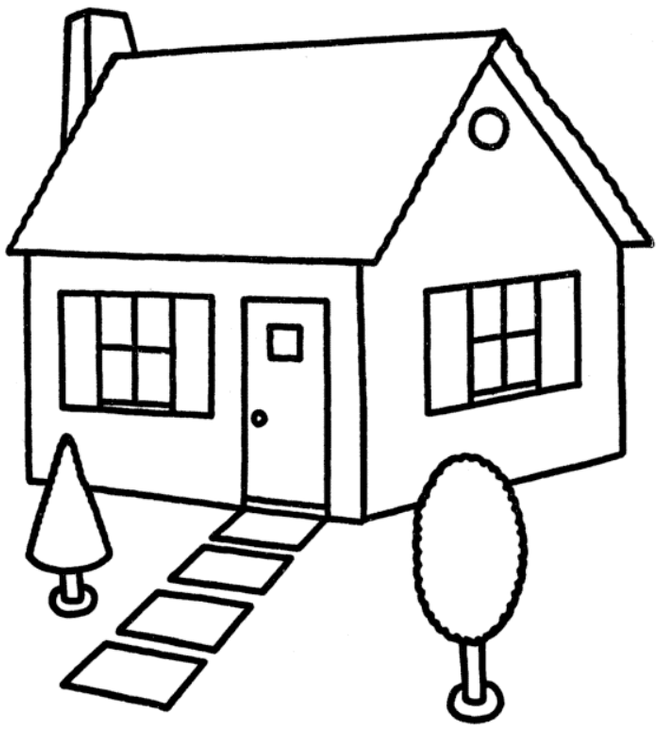
digipres



There’s a problem

Checksum verification analogy

digipres



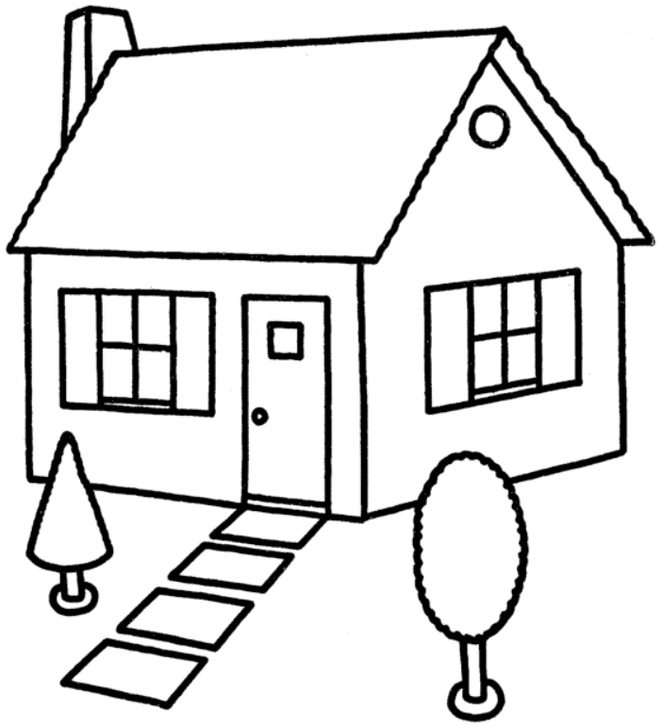
There's a problem

av archiving

There's a problem

Checksum verification analogy

digipres



There's a problem

av archiving

There's a problem

av archiving
with framemd5

There's a problem

framemd5

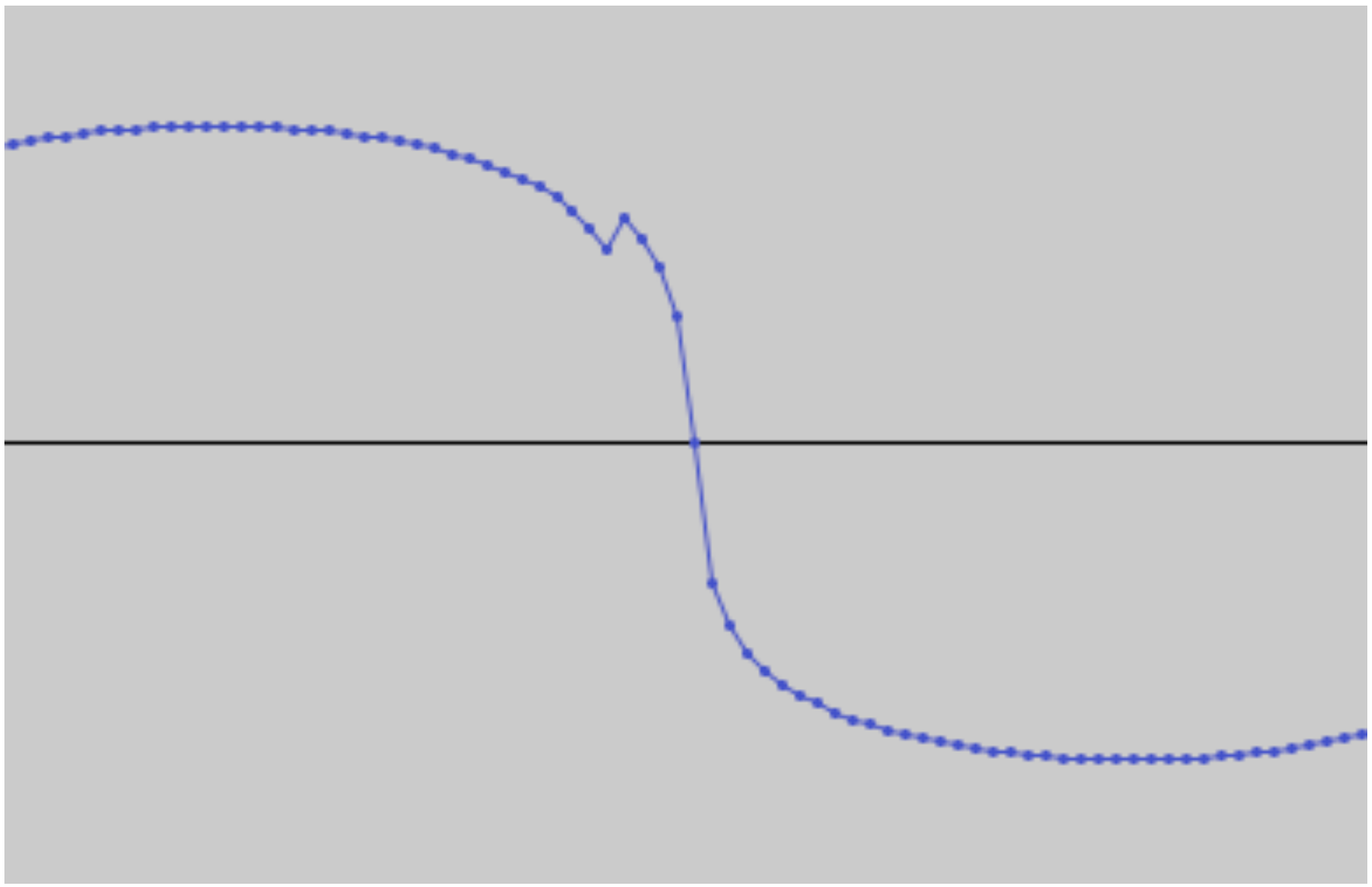
stream_index, packet_dts, packet_pts, packet_duration, packet_size, MD5

```
#tb 0: 1001/30000
0,          0,          0,        1,  1228800, 39cc3d1589ea601881abddd07fc8a32b
0,          1,          1,        1,  1228800, bbdbd60d757017ae0f75a505e589f5dc
0,          2,          2,        1,  1228800, 94b7c46fb0d456209fca3206bb3f20f2
0,          4,          4,        1,  1228800, b4a2e57b784164716a7f4f5c4b1fa22e
0,          5,          5,        1,  1228800, 1d16a3d98d9b2667419ad84814d9c74e
0,          6,          6,        1,  1228800, 28bdf2224aac89bb445d3e55755792bb
0,          7,          7,        1,  1228800, 4b15f8b8f27e94aa515c76b83fa566ef
0,          8,          8,        1,  1228800, a544793f6bb30d97ccf43549e74c7748
0,         10,         10,        1,  1228800, f222da8e7be0dfba8f6f70fbbaf82f47
0,         11,         11,        1,  1228800, bc42b9f41147d4db582ce91f3ce641b3
0,         12,         12,        1,  1228800, dbc1bd60575626e3f46c6726b632e8be

#tb 0: 1001/30000
0,          0,          0,        1,  1228800, 39cc3d1589ea601881abddd07fc8a32b
0,          1,          1,        1,  1228800, bbdbd60d757017ae0f75a505e589f5dc
0,          2,          2,        1,  1228800, 94b7c46fb0d456209fca3206bb3f20f2
0,          4,          4,        1,  1228800, b4a2e57b784164716a7f4f5c4b1fa22e
0,          5,          5,        1,  1228800, 1d16a3d98d9b2667419ad84814d9c74e
0,          6,          6,        1,  1228800, 35f67d5a7dd9eab6938a38b56d785648
0,          7,          7,        1,  1228800, 4b15f8b8f27e94aa515c76b83fa566ef
0,          8,          8,        1,  1228800, a544793f6bb30d97ccf43549e74c7748
0,         10,         10,        1,  1228800, f222da8e7be0dfba8f6f70fbbaf82f47
0,         11,         11,        1,  1228800, bc42b9f41147d4db582ce91f3ce641b3
0,         12,         12,        1,  1228800, dbc1bd60575626e3f46c6726b632e8be
```

FLAC are internally verified by md5 per stream and crc per audio frame

```
ffmpeg -v warning -i sine.flac -f null -
[flac @ 0x7fd0cb05e800] CRC error at PTS 96768
```



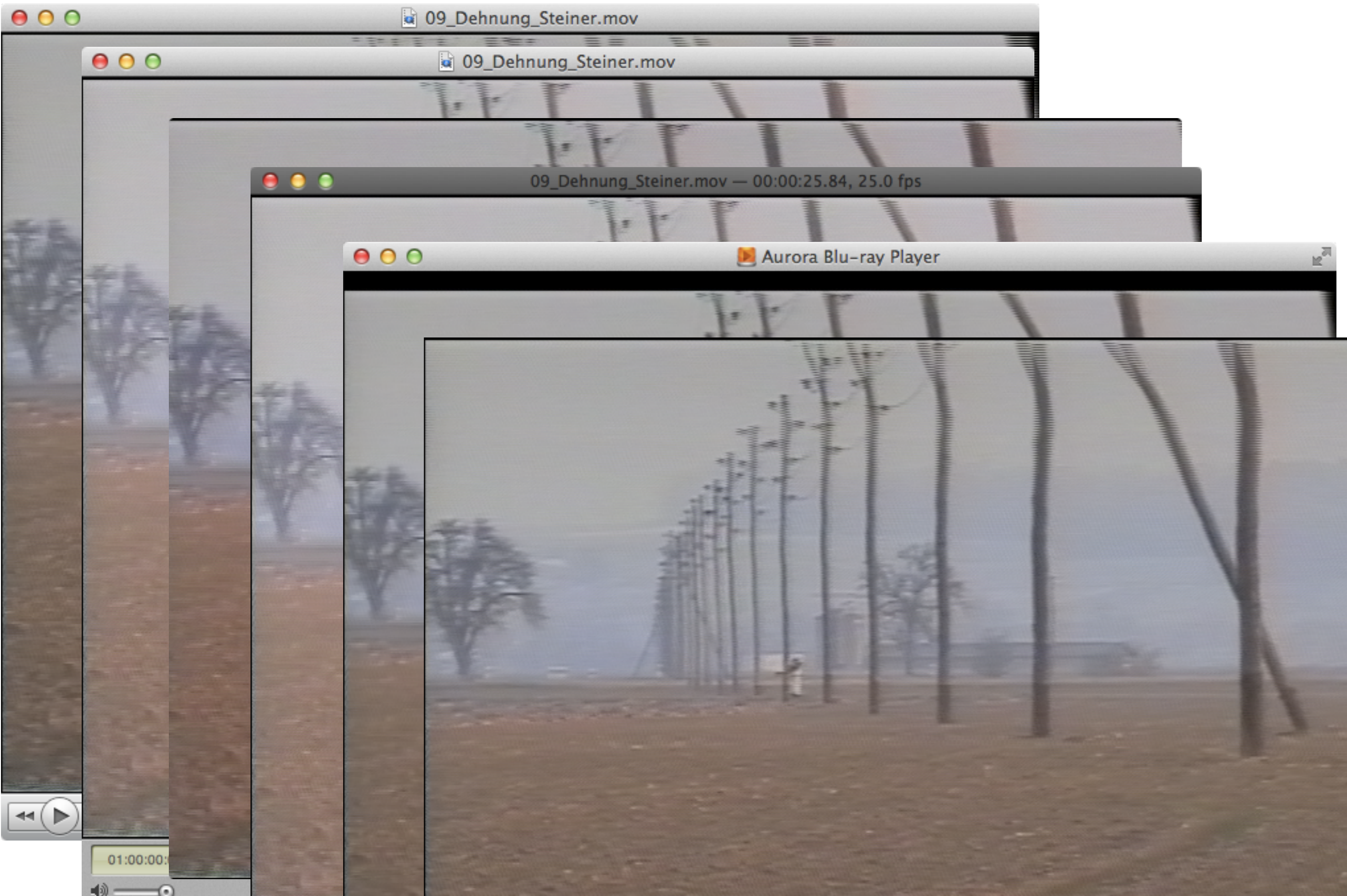
“[ffv1 @ 0x7f9855046e00] CRC mismatch FC686A4F! frame 215”

Significant Characteristics

Frame Size
Frame Rate
Color Space: Color Matrix
Luma Range: Broadcast or Full
Aspect Ratio
Interlacement
Duration / Timeline
Audio Channel Configuration / Arragement

```
264  265
     266  +static const char *j2ki_field_rate_companies[] = { "Canopus", "SAMMA" };
     267  +
265  268   static int64_t klv_decode_ber_length(AVIOContext *pb)
266  269   {
267  270       uint64_t size = avio_r8(pb);
@@ -1331,6 +1334,7 @@ static int mxf_add_timecode_metadata(AVDictionary **pm, const char *key, AVTimec
1331 1334
1332 1335   static int mxf_parse_structural_metadata(MXFContext *mxf)
1333 1336   {
     1337  +    AVFormatContext *s = mxf->fc;
1334 1338       MXFPackage *material_package = NULL;
1335 1339       MXFPackage *temp_package = NULL;
1336 1340       int i, j, k, ret;
@@ -1553,8 +1557,24 @@ static int mxf_parse_structural_metadata(MXFContext *mxf)
1553 1557                   case AV_CODEC_ID_JPEG2000:
1554 1558                       if (descriptor->frame_layout == SegmentedFrame ||
1555 1559                           descriptor->frame_layout == SeparateFields) {
1556      -                        st->codec->time_base = st->time_base;
1557      -                        st->time_base = (AVRational) { st->time_base.num, st->time_base.den * 2};
     1560  +                        int l, field_rate = 0;
     1561  +                        AVDictionaryEntry *entry = av_dict_get(s->metadata, "company_name", NULL, 0);
     1562  +                        if (entry) {
     1563  +                            for (l = 0; l < FF_ARRAY_ELEMS(j2ki_field_rate_companies); l++) {
     1564  +                                if (av_stristr(entry->value, j2ki_field_rate_companies[l])) {
     1565  +                                    av_log(s, AV_LOG_INFO, "J2ki sample rate will be interpreted as field rate for company: %s
     1566  +                                    field_rate = 1;
     1567  +                                    break;
     1568  +                                }
     1569  +                            }
     1570  +                        }
     1571  +
```

Sustaining Consistent Presentation



Username

Password

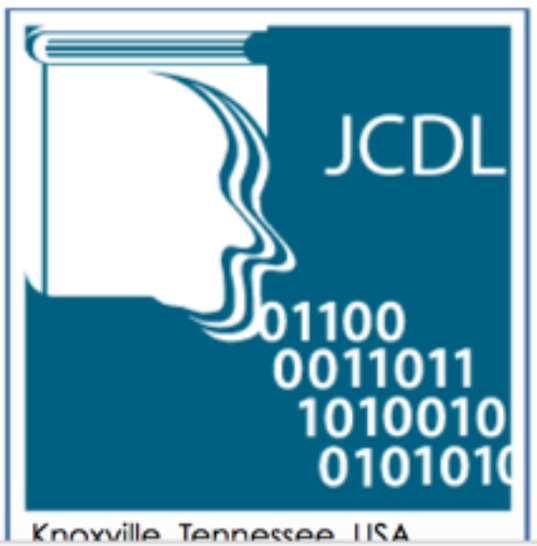
login

MEDIA PARTNER

DIGITAL CULTURE .net

Search

RELEVANT NEWS from Digitalmeetsculture



PREFORMA, FUTURE MEMORY STANDARDS

PREFORMA

Overview

Memory institutions are facing increasing transfers of electronic documents and other media content for **long term preservation**. Preservation models are often inspired by ISO 14721:2003, known as "the OAIS model", where transfers and preservation are built on information packages containing both data and metadata.

Data are normally stored in specific file formats for documents, images, sound, video etc. that are produced by software from different vendors. Even if the transferred files are in standard formats, the implementation of **standards** cannot be guaranteed. The software implementing standards for the production of the electronic files is not in control neither by the institutions that produces them nor by the memory institutions. Conformance tests of transfers are done, but are not totally reliable. This poses problems in long-term preservation. Data objects meant for preservation, passing through an uncontrolled generative process, can jeopardise the whole preservation exercise.

The overall intention of **PREFORMA** project (PREservation FORMAts for culture information/e-archives) is to research critical factors in the quality of standard implementation in order to establish a long-term sustainable ecosystem around developed tools with a variety of stakeholder groups. The tools should be innovative and provide a reference implementation of the most common file format standards for the assessment of the collections to be archived and for the

COORDINATOR

Riksarkivet

TECHNICAL COORD.

PROMOTER

PARTNERS

PACKED
Centre d'Expertise pour le Patrimoine Numérique

Fraunhofer IDMT

Enabling Preservation by means of Open Source

dave rice
@dericed
#fosdem
2015-01-31